Health Policy and Planning, 2020, 1–9
doi: 10.1093/heapol/czaa001
Original Article

OXFORD

No effects of pilot performance-based intervention implementation and withdrawal on the coverage of maternal and child health services in the Koulikoro region, Mali: an interrupted time series analysis

David Zombré[1,*], Manuela De Allegri[2] and Valéry Ridde[1,3]

[1]Department of Social and Preventive Medicine, University of Montreal Public Health Research Institute – IRSPUM, Pavillon 7101 avenue du Parc, C.P 6128 Succursale C, Local 3224, Montréal, Québec H3C 3J7, Canada, [2]Heidelberg Institute of Global Health (HIGH), Medical Faculty and University Hospital, University of Heidelberg, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany, and [3]RD (French Institute for Research on sustainable Development), CEPED (IRD-Université Paris Descartes), Universités Paris Sorbonne Cités, ERL INSERM SAGESUD, 45 rue des Saints-Pères 75006 Paris, France

*Corresponding author. Department of Social and Preventive Medicine, University of Montreal Public Health Research Institute – IRSPUM, Pavillon 7101 avenue du Parc, C.P 6128 Succursale C, local 3224, Montréal, Québec H3C 3J7, Canada. E-mail: david.zombre@umontreal.ca

Accepted on 2 January 2020

Abstract

Performance-based financing (PBF) has been promoted and increasingly implemented across low- and middle-income countries to increase the utilization and quality of primary health care. However, the evidence of the impact of PBF is mixed and varies substantially across settings. Thus, further rigorous investigation is needed to be able to draw broader conclusions about the effects of this health financing reform. We examined the effects of the implementation and subsequent withdrawal of the PBF pilot programme in the Koulikoro region of Mali on a range of relevant maternal and child health indicators targeted by the programme. We relied on a control interrupted time series design to examine the trend in maternal and child health service utilization rates prior to the PBF intervention, during its implementation and after its withdrawal in 26 intervention health centres. The results for these 26 intervention centres were compared with those for 95 control health centres, with an observation window that covered 27 quarters. Using a mixed-effects negative binomial model combined with a linear spline regression model and covariates adjustment, we found that neither the introduction nor the withdrawal of the pilot PBF programme bore a significant impact in the trend of maternal and child health service use indicators in the Koulikoro region of Mali. The absence of significant effects in the health facilities could be explained by the context, by the weaknesses in the intervention design and by the causal hypothesis and implementation. Further inquiry is required in order to provide policymakers and practitioners with vital information about the lack of effects detected by our quantitative analysis.

Keywords: Performance-based financing, health services coverage, policy evaluation, interrupted time series, Mali

Key Messages

- Performance-based financing (PBF) has generally been unable to produce changes in maternal and child health indicators in Mali.
- The absence of change attributable to PBF could be explained by weaknesses in the design of the specific programme under analysis, the very short duration of its implementation and a weak implementation process.
- A clear programme theory of change would ensure both a well-thought-through programme design and evaluations that explain the mechanism of change.

Introduction

Since maternal and child health remain a priority in low-income countries, all possibilities continue to be explored to improve access to healthcare with the aim of reducing the disease burden on mothers and their children (Lassi *et al.*, 2016). Thus, performance-based financing (PBF) has been promoted and increasingly implemented to increase the utilization and quality of primary healthcare, often with a specific focus on maternal and child health.

Findings from studies conducted in Africa on the effectiveness of PBF on maternal and health outcomes and quality of care, including evidence syntheses and reviews, are generally mixed (Witter *et al.*, 2012; Das *et al.*, 2016; Renmans *et al.*, 2016; Paul *et al.*, 2018) and uncertain (Wiysonge *et al.*, 2017) so that no general conclusions can currently be drawn on the effectiveness of PBF (Witter *et al.*, 2012), with some authors even urging a reconsideration of the value of introducing PBF in low-income countries (Paul *et al.*, 2018). For example, in Rwanda, the introduction of PBF resulted in a moderate increase in assisted deliveries (Basinga *et al.*, 2011). By considering the effects on service quality, user costs and equity, the evaluation of the PBF scheme in Tanzania revealed positive effects on the coverage of institutional deliveries and the provision of two doses of antimalarial during pregnancy, while no overall effect on the use of non-targeted services was found (Binyaruka *et al.*, 2015). However, no equity effects were identified and no effect was found in terms of improvement in antenatal content of care and patient satisfaction with inter-personal care for targeted services and financial protection (Binyaruka *et al.*, 2015). In Burkina Faso, a study showed that the pilot PBF programme in three districts resulted in an increase in the average number of curative visits, deliveries and postnatal visits by 27.7%, 9.2% and 9.9%, respectively (Steenland *et al.*, 2017), while a more recent study conducted in the same context on the large pilot aimed at partial scale-up concluded that the programme did not bear clear effects on maternal and child health indicators half-way through its implementation life (Zizien *et al.*, 2018). A similar conclusion had already been drawn in the Democratic Republic of Congo and in Uganda, where PBF did not significantly change patterns of health service utilization (Ssengooba *et al.*, 2012; Huillery and Seban, 2014). However, a household survey also conducted in the Democratic Republic of Congo found statistically significant improvements in the quality of services regarding the availability of medicines, the perceived quality of care, the hygiene of health facilities and the level of respect shown at the reception desk (Zeng *et al.*, 2018).

Between 2012 and 2013 Mali, in collaboration with the Dutch Cooperation (SNV), implemented a PBF pilot programme in Community Health Centre (CSCOMs) from three health districts in the Koulikoro region. This pilot programme was rolled out over 16 months and then discontinued in December 2013. An evaluation carried out by the pilot PBF programme implementation team showed an increase in the use of maternal and child health services and a significant improvement in the quality of care and revenue generation capacity of community health associations (ASACOs) (Toonen *et al.*, 2014). The existing evaluation suffers from an important limitation, since it estimated effects without considering a control group, compromising the study's internal validity (Gautier, 2016). In addition, the existing evaluation did not even attempt to estimate the effects of programme withdrawal on relevant outcomes.

Overall, given that impact evaluations of PBF programmes have consistently reported mixed results (Ssengooba *et al.*, 2012; Witter *et al.*, 2012; Renmans *et al.*, 2016; Steenland *et al.*, 2017), existing reviews have been called for more investigation (Witter *et al.*, 2012).

The pilot PBF intervention in Mali was implemented sequentially in several CSCOMs and then withdrawn 16 months later. Hence, the availability of Health Management Information Systems (HMIS) data systematically collected before, during and after the pilot phase of the PBF offers a unique opportunity to thoroughly evaluate its impact beyond the valuable, yet simpler, analysis already completed by the implementation team.

This is the first study to examine the effects of the implementation and subsequent withdrawal of the PBF pilot programme in the Koulikoro region of Mali in the trends change of relevant MNCH indicators targeted by the programme. In addition, we want to demonstrate the relevance of combining time series analysis relying on the use of HMIS data with a pragmatic design and analytical approach in order to draw valid conclusions about the impact of complex health interventions (Wagenaar *et al.*, 2015).

Methods

Settings

This study was conducted in Koulikoro, the second largest region in the country, located 60 km from Bamako in north-eastern Mali (CAD-Mali, 2011). The region is dominated by plateaus and plains, and 130 km of it is crossed by the Niger River. Based upon the Bamako initiative model, the health system is based on a network of CSCOMs run by Community Health Associations (ASACOs) made up of representatives of the population (Ponsar *et al.*, 2011). They constitute the front-line services, offering a minimum package of essential care while referring the most serious disease cases to the reference health centre (CSREFs) (Ponsar *et al.*, 2011). The Ministry of Health (MoH) builds the centres, equips them and supplies the initial stocks of medicines, and it allocates and remunerates the responsible health worker(s) (Ponsar *et al.*, 2011). Some additional

Table 1 Outcomes and their measurement

Outcome variable	Measurement
Proportion of under-five curative consultations	Number of curative consultations/ number of children under five in the catchment area
Proportion of fully immunized children	Number of children who received the measles vaccine (VAR)/ number of children aged 0–11 months
Proportion of assisted deliveries	Number of assisted deliveries/ numbers of expected deliveries
Proportion of postnatal consultations	Number of post-natal consultations/number of assisted deliveries
Proportion of women receiving a vitamin A dose	Number of women who received a dose of vitamin A after delivery/ number of assisted deliveries

expenses of CSCOMs are subsidized by the government and by the multiple non-governmental organizations (NGOs), but a substantial portion of total health expenditure (62%) is still borne by households through out-of-pocket expenditure (World Health Organization, 2014).

Intervention

With the ambition of tackling maternal and neonatal mortality, the MoH rolled out a pilot PBF programme in 2012. This pilot project, financed by the SNV, was meant to be a 'Malian PBF', i.e. it was meant to account for the Malian-specific institutional and administrative context of decentralization in the health sector (Toonen *et al.*, 2014). Performance contracts were signed with the 26 CSCOMs management committees, whereby facilities were incentivized and received additional financial compensation based on the achievement of a set of pre-defined quantity and quality indicators related to maternal and child health outcomes (Toonen *et al.*, 2014). PBF indicators included under-five curative consultations, prenatal and postnatal consultations, children immunization, family planning, assisted deliveries, number of children or women referred to the SCCOMs, and women receiving a vitamin A dose.

Performance scores were calculated, and payments were effective for CSCOMs that reached at least 75% of the target relative to both quantity and quality of service provision. Full payment was granted to CSCOMs that reached 100% of the target. Verification of both quantity and quality performance indicators was entrusted to the State Technical Services, while an independent counter-verification was carried out by a local NGO. Sixty per cent of the funds generated through PBF was to be re-invested in the facility and 40% used to pay staff bonuses. The MoH expected the PBF programme to result in increased service delivery and improved quality of care by improving the working conditions of health personnel as well as their motivation. PBF was restricted to the 26 CSCOMs in the Banamba, Fana and Dioila districts that benefited from the intervention; the remaining 95 health facilities in Koulikoro were excluded from the intervention.

Districts were selected by SNV NGO-based prior performances on governance and on a set of service use indicators (Toonen *et al.*, 2014). Given the human resource constraints within the implementing NGO team, the programme was implemented sequentially in districts, meaning that every month an additional three CSCOMs were included in the intervention. The idea was to allow the intervention team to build on the experience of the first CSCOMs to correct imperfections in the course of its work. The pilot PBF intervention was withdrawn in December 2013 when the SNV funds stopped at the end of its implementation because of financial constraints.

Study design

We relied on a controlled interrupted time series design (Lopez Bernal *et al.*, 2018) and applied two interruptions. The first indicating the PBF introduction (specific to the single facilities) and the second indicating the intervention withdrawal. The observation window covered the period from the first quarter of 2009 to the last quarter of 2015, resulting in 15 quarters before the beginning of the intervention, 6 quarters during the intervention implementation and 7 quarters after the withdrawal of the intervention. Due to pragmatic implementation considerations, CSCOMs were included in the intervention at different points in time: seven CSCOMs received PBF in the 15th quarter; one started in the 17th quarter; five started in the 18th quarter; seven in the 19th quarter; and six in the 20th quarter. The intervention was withdrawn in the 21st quarter in all the CSCOMS.

Data sources

Since data are compiled quarterly, for both intervention and control CSCOMs in the Koulikoro region, we extracted relevant data on maternal and child health service delivery from the HMIS to cover the period from the first quarter of 2009 to the fourth quarter of 2015 to form a 27-month time series. In addition, from the same source and over the same period, we obtained data on the number and qualification of health workers per facility, the target population in the relevant catchment areas and its geographical distribution according to distance. The reliability and validity of HMIS data in the context of Mali have been proven in previous studies (Ponsar *et al.*, 2011; Heinmüller *et al.*, 2012).

Outcome variables and their measurement

We selected our outcome variables to reflect the health service indicators included in the programme and by considering which data could be extracted from the HMIS. Table 1 summarizes the outcome variables included in our analysis and their measurement. For each indicator, and specific to each quarter, we generated relevant proportions by dividing the absolute count of events (e.g. number of under-five curative consultations) by the size of the specific target population (e.g. the total number of children under five residing in the catchment area).

Statistical analyses

Model selection

In evaluating public health intervention using longitudinal data, threats to internal validity can be limited if the intervention and the control group are comparable both in terms of pre-intervention covariates and in terms of the level and trend of change in the outcomes of interest being observed (Lee and Little, 2017; Handley *et al.*, 2018). In our case, this would mean that prior to the PBF launch, the CSCOMs included in our analysis should not differ substantially in terms of their facility characteristics nor in terms of the level and trend of change in maternal and child health service utilization. To limit this threat, we used an analytical strategy that compared results from a controlled interrupted time series design (Lopez Bernal *et al.*, 2018) combined with a covariates adjustment with results from a propensity score weighting method (McCaffrey *et al.*, 2004) in order to choose the one that provided good fits to the data.

Propensity score weighting

In practice, the propensity score represents the conditional probability of a PBF CSCOM becoming treated when matched over a range of observable characteristics (Rosenbaum and Rubin, 1983). The propensity score weighting approach follows a three-step approach (Linden and Adams, 2010; Lee and Little, 2017). First, we estimated the PS using GBM algorithm that optimized balance for the purposes of estimating the ATT parameter. Second, weights are constructed based on the propensity score and treatment assignment. Third, these weights are then used within a regression framework (Linden and Adams, 2010; Lee and Little, 2017) to provide the intervention effect in the trend in maternal and child health service utilization rates.

To generate the propensity scores, we included two types of variables in our model: (1) contextual variables related to the proportion of the population of the CSCOM catchment area living within 5 km and (2) initial performance variables, i.e. average performance of CSCOMs related to the level of the use of maternal and child health 2 years before the start of the intervention (see Supplementary Annexe SA for more information). Given that the propensity score method per se is not the focus of this article, details on the model estimation are provided in the Supplementary Annexe SC.

Covariate adjustment

In the model with covariates adjustment, we included three time-invariant covariates. The first variable is a contextual variable describing the accessibility to services defined as the proportion of the population that lived <5 km from each health centre. The second variable represents a health service-related factor measuring the health workforce density defined as the number of health workers (nurses, midwives and other health professionals) per 1000 children under five. The third variable describes the initial performance, i.e. average performance of CSCOMs related to the level of the use of maternal and child health 2 years before the start of the intervention (see Supplementary Annexe SA for details on the calculation).

Choice of the best-fitting model

To determine the best-fitting multilevel model, we compared the GBM propensity weighting model with a simple multilevel model, adjusting for three covariates. We used the Akaike information criterion (AIC) (Akaike, 1998) for this purpose. Afterwards, we relied on the Wald test to compare the quarterly trend differences before, during PBF implementation and after its withdrawal.

Justification of the modelling strategy

As the data have a multilevel structure, made by quarterly measures nested within CSCOMs, we evaluated the effects of the pilot PBF programme and its removal using multilevel negative binomial modelling, controlling for secular trend, over-dispersion (Jung *et al.*, 2006) and confounding by season, by treating each quarter as a separate category with a fixed dummy parameter to capture the magnitudes of seasonality and to control for seasonal patterns (Barnett and Dobson, 2010).

Our analytical strategy was based on the following consideration: first, compared with demand-side intervention as user fee removal that would be expected to follow a step change model (Lopez Bernal *et al.*, 2018) with the outcomes that could respond rapidly (Zombré *et al.*, 2017), we have assumed that the PBF is an intervention that acts directly on the supply of healthcare. Its effect on the demand for healthcare is necessary through the motivation of health workers (Renmans *et al.*, 2017) and the improvement of the quality of care (Eichler, 2006). This will require a delay before observing a change in the demand for care (Eichler, 2006; Renmans *et al.*, 2017). The impact of PBF on the health indicators we have measured should occur gradually with a latency period that would exclude an immediate effect. This was confirmed by the visual analysis and the sensitivity analysis that we carried out with the multiple basic level estimate exclusively in the 26 CSCOMs of intervention. It is for these reasons that we choose to analyse the impact of the intervention in the trend in the health service utilization.

Given that the difference-in-differences method (Dimick and Ryan, 2014) and the synthetic control approach (Abadie *et al.*, 2010) are not the most appropriate modelling solutions to respond to our research objectives, we used the linear spline regression method (Lawrence, 2002). The linear spline regression method allows a subtle change in the trend in response to a possible effect of the pilot PBF programme or its withdrawal to examine the trend in maternal and child health service utilization rates prior to the pilot PBF intervention during its implementation and after its withdrawal in the intervention CSCOMs compared with the control CSCOMs.

The basic model for the multilevel negative binomial regression with linear trend assumption is defined as follows:

$$\log(U_{it}) = \beta_{i0} + \beta_1\mathrm{Group} + \beta_2\mathrm{Time} + \beta_3\mathrm{Group}*\mathrm{Time} + \beta_4\mathrm{PBF}_{\mathrm{time}} + \beta_5\mathrm{Group}_{\mathrm{PBF}}\mathrm{time} + \beta_6\mathrm{PostPBF}_{\mathrm{time}} + \beta_7\mathrm{Group}_{\mathrm{PostPBF}}\mathrm{Time} + \beta_8\mathrm{Health}_{\mathrm{Worker}}i + \beta_9\mathrm{Distance}_i + \beta_{10}\mathrm{Performance}_i + \beta_{11}\mathrm{Quarter}_2 + \beta_{12}\mathrm{Quarter}_2 + \beta_{13}\mathrm{Quarter}_3 + \beta_{14}\mathrm{Quarter}_4 + \nu_{0i} + e_{it}.$$

U_{it} represents the proportion of each of the relevant outcome expected in CSCOM *i* each quarter *t*. These values are assumed to follow a negative binomial distribution. The exp factor $\mathrm{Exp}(\beta_{i0})$ is used to estimate the initial level utilization for each indicator in control CSCOMs, while factor $\mathrm{Exp}(\beta_{i0}+\beta_1)$ allows us to define these same initial proportions in the intervention CSCOMs. The variable *Group* takes 1 value for PBF CSCOMs and 0 otherwise. The terms $\mathrm{Exp}(\beta_2 + \beta_3)$ and $\mathrm{Exp}(\beta_2)$ represent the trend in the utilization rates for the relevant outcome indicators in both the intervention and the control CSCOMs before the intervention onset. The terms $\mathrm{Exp}(\beta_4 + \beta_5)$ and $\mathrm{Exp}(\beta_4)$ allow the estimation of the trends in the intervention and control groups during the programme implementation phase. $\mathrm{Exp}(\beta_6 + \beta_7)$ and $\mathrm{Exp}(\beta_6)$ made it possible to estimate the trend of utilization for the relevant outcomes after withdrawal of the pilot PBF programme in the intervention and control groups. β_8, β_9 and β_{10} allow to adjust for the effect of the covariates related to the density of health workers, the distribution of the population and the initial performance of each health centre in terms of curative consultations, assisted deliveries, immunization, administration of vitamin A, and postnatal consultations. Finally, we treated each quarter as a separate dummy variable with a fixed parameter to capture the magnitude of seasonality and to control for seasonal patterns (Barnett and Dobson, 2010). We included a random term (ν_{0i}) to allow the initial level of service utilization to vary between CSCOMs. However, because of convergences issues, we have not been able to include random slope effects in order to examine possible heterogeneity across intervention facilities. We performed all the analyses using Intercooled Stata version 15 software (Stata Corp., College Station, TX, USA). Trend and differences in trend and their confidence intervals were obtained using the command nlcom of Stata, which computes standard errors and confidence intervals for non-linear combinations of estimate parameters using the delta method (Feiveson, 1999).

Results

When comparing the values of the AIC statistics, results from the multilevel regression models show that the simple multilevel model adjusting for the time-invariant covariates provided a better fit for all our outcome variables than the GBM propensity weighting model (see details on regression models at Supplementary Annexe SE). Therefore, we assessed the impact of the pilot PBF and its withdrawal on maternal and child health outcomes adjusted for health workers density, dispersion of population, the initial performance in term of maternal and child healthcare service and seasonal trend. The results of the two modelling approaches are presented for robustness check (Table 2 and Table S3 in Supplementary Annexe SE).

Effects of the pilot PBF programme on maternal and child health indicators

As shown in Table 2 and Figure 1, the initial rate of maternal and child health indicators targeted by PBF was almost similar between the two groups ($P > 0.10$).

Over the period prior to the introduction of the pilot PBF programme, the trend was downward and relatively stable in the two groups for assisted deliveries (trend difference = 0.01, $P > 0.45$) and proportion of fully immunized children (trend difference = 0.01, $P > 0.45$) and was increasing almost in a similar pace in both groups for postnatal visits (trend difference = 0.01, $P > 0.28$), women receiving a vitamin A dose (trend difference =0.01, $P = 0.48$) and for under-five curative consultations (trend difference = 0.01, $P > 0.20$).

During the implementation of the pilot PBF, the pace of growth for the maternal and child health outcomes was almost similar in both groups to that of the period prior to the introduction of the pilot PBF programme for the postnatal visits ($P > 0.98$), under-five consultations ($P > 0.93$) and fully immunized children ($P > 0.57$). However, our results showed a statistically non-significant decrease of the proportion of assisted deliveries (trend difference = −0.04, $P = 0.1$). Thus, we found no difference in the trend of the use of the maternal and child health services ($P > 0.10$), suggesting that the pilot PBF did not impact the trend of the targeted maternal and child health outcomes in the intervention CSCOMs.

Finally, after withdrawal of the pilot PBF programme, we observed a statistically non-significant decrease of the trend of postnatal visits (trend difference = −0.03, $P = 0.44$), of the proportion of women receiving vitamin A (trend difference = −0.02, $P = 0.59$), under-five curative consultations (trend difference = −0.01, $P = 0.87$) and fully immunized children (trend difference = −0.01, $P = 0.58$) in intervention group compared with the control group. In the same lines, our results showed a statistically non-significant increase of the trend of the proportion of assisted deliveries (trend difference = 0.05, $P = 0.10$). In sum, we found no statistically significant difference in the trend of the use of maternal and child health services ($P > 0.10$), suggesting that the removal of the intervention did not impact the trend of the targeted maternal and child health outcomes in the intervention CSCOMs.

Discussion

Using pragmatic analytical approaches, this study contributes to the ongoing debate on the effectiveness of PBF to improve maternal and child health outcomes, specifically looking at changes in health service utilization patterns during PBF implementation and after its withdrawal. In contrast to the results of previous studies conducted in other contexts (Basinga *et al.*, 2011; Bonfrer *et al.*, 2014; Steenland *et al.*, 2017), neither the introduction nor the withdrawal of the pilot PBF programme globally bore a significant impact in the trend of maternal and child health services utilization in the Koulikoro region of Mali.

While our study did not find any evidence of an impact of the pilot PBF intervention on relevant outcome indicators, a previous study on the sustainability of the pilot PBF in our study context, suggested that several gains were generated by the pilot PBF programme in Mali, particularly in terms of long-term investments in human and material resources, the integration of various tasks and procedures and the creation of trust between stakeholders (Seppey *et al.*, 2017). However, the absence of significant global effects of the PBF programme could be explained by other important factors related to the context of the PBF, its design and the durability of its implementation. Indeed, the pilot PBF programme was a very new approach in Mali, and the implementation team did not receive sufficient financial and human resources to start the programme simultaneously in the 26 health centres. In addition, the programme was rolled out in phases with delays taking place in the distribution of bonuses to health workers during implementation of the pilot PBF programme (Abdourahmane *et al.*, 2018). These delays compromised to some extent health workers' motivation as well as the ability to put into effect investments towards quality of care improvement. As a consequence, they probably contributed to dilutions or constraints on the average effect of the programme across all CSCOMs that experienced it. Furthermore, PBFs have been deployed and removed in a context where 62% of health expenditure is borne by households through out-of-pocket expenditure. So, even if the quality of healthcare may have been improved with the intervention and did not change with its withdrawal, and intervention had sufficient time to be deployed, maternal and child service use would be tough to change with these high user fees. This reinforces our idea that in this context where 62% of health expenditure is borne by the household, the healthcare demand would be more sensitive to interventions that improve the affordability of healthcare rather than interventions that improve the motivation of health workers and the quality of care.

In addition, while the implementation of public health interventions requires a certain amount of time to produce the expected effects, the very short duration of the pilot PBF programme in Mali (16 months) did not allow the project team to draw on the experience of the first health centres to gradually correct imperfections as they were planned (Toonen *et al.*, 2014). It is also possible that the pilot PBF intervention did not have sufficient time to drive demand at the population level, and the variability in the implementation of one CSCOM from another also made it difficult to produce positive effects in all CSCOMs. In light of our findings, it is plausible to assume that a subsequent PBF project supported by the World Bank starting in 2017 in 10 of the region's districts will lead to similar results, due to the fact that implementation delays ultimately constrained its actual implementation to only 8 months (Abdourahmane *et al.*, 2018). From the perspective of evaluations producing useful results in decision-making in public health (Wholey, 2004), this leads us to question the evaluability of PBF intervention (D'Ostie-Racine *et al.*, 2013) and the appropriateness of programme implementation strategy in the context of low and middle income countries (LMIC). Thus, while all PBF programmes had weak theories of change at start-up, the need for a good

Table 2 Effect of the pilot PBF programme and its withdrawal on the trend of maternal and child outcomes in Koulikoro region (Mali)

Parameter	Assisted deliveries		Postnatal Consultations		Postpartum vitamin A		Under-five consultations		Fully immunized children	
	Simple multilevel	Multilevel + PSW	Simple multilevel	Multilevel + PSW	Simple multilevel	Multilevel + PSW	Simple multilevel	Multilevel + PSW	Simple multilevel	Multilevel + PSW
Initial rate										
Control group	32.13 [23.49 – 40.77]	70.13 [59.47 – 80.8]	84.69 [58.04 – 111.35]	53.59 [41.93 – 65.25]	24.19 [16.92 – 31.47]	42.45 [37.11 – 47.78]	141.02 [92.8 – 189.25]	247.08 [204.29 – 289.87]	90.12 [77.69 – 102.54]	95.1 [87.43 – 102.77]
Intervention group	31.85 [26.22 – 37.47]	68.5 [44.61 – 92.38]	76.87 [52.27 – 101.47]	58.56 [44.89 – 72.23]	26.3 [19.73 – 32.87]	36.73 [27.54 – 45.92]	172.54 [132.97 – 212.1]	184.19 [135.41 – 232.96]	81.56 [72.16 – 90.96]	102.82 [86.7 – 118.95]
Difference	0.28 [−7.08 – 7.65]	−1.64 [−27.7 – 24.43]	7.82 [−16.8 – 32.44]	4.97 [−12.99 – 22.93]	−2.1 [−8.73 – 4.52]	−5.71 [−16.35 – 4.93]	−31.52 [−75.88 – 12.85]	−62.89 [−127.47 – 1.68]	8.55 [−3.79 – 20.9]	7.72 [−10.31 – 25.75]
Pre-PBF trend										
Control group	0.98 [0.97 – 0.99]	0.98 [0.98 – 0.99]	1.02 [1 – 1.03]	1.01 [0.99 – 1.02]	1.02 [1 – 1.04]	1.02 [1.01 – 1.02]	1.03 [1.01 – 1.05]	1.02 [1.01 – 1.03]	1 [0.98 – 1.01]	1 [0.99 – 1]
Intervention group	0.99 [0.97 – 1.02]	0.99 [0.97 – 1.02]	1.01 [0.99 – 1.02]	1.02 [1 – 1.04]	1.01 [1 – 1.02]	1.03 [1.01 – 1.05]	1.02 [1.01 – 1.02]	1.03 [1.01 – 1.05]	1 [0.99 – 1.01]	0.99 [0.98 – 1.01]
Difference	−0.01 [−0.02 – 0.03]	0.01 [−0.02 – 0.04]	0.01 [−0.01 – 0.03]	0.02 [−0.01 – 0.04]	0.01 [−0.01 – 0.03]	0.01 [−0.01 – 0.03]	0.01 [−0.01 – 0.04]	0.01 [−0.01 – 0.04]	0 [−0.02 – 0.01]	−0.01 [−0.02 – 0.01]
Trend during PBF rollout										
Control group	0.98 [0.93 – 1.03]	1.02 [1 – 1.04]	1.01 [0.96 – 1.06]	1.01 [0.97 – 1.04]	0.96 [0.91 – 1]	0.94 [0.92 – 0.97]	1.01 [0.96 – 1.06]	1.01 [0.99 – 1.03]	1 [0.98 – 1.03]	1 [0.98 – 1.01]
Intervention group	1.02 [1 – 1.04]	0.98 [0.93 – 1.02]	1.01 [0.97 – 1.05]	1.01 [0.96 – 1.06]	0.95 [0.92 – 0.97]	0.95 [0.9 – 0.99]	1.01 [0.99 – 1.03]	1.01 [0.96 – 1.06]	0.99 [0.98 – 1.01]	1.02 [0.99 – 1.05]
Difference	−0.04 [−0.09 – 0.01]	−0.04 [−0.09 – 0.01]	0 [−0.06 – 0.06]	0 [−0.06 – 0.06]	0.01 [−0.04 – 0.06]	0 [−0.05 – 0.06]	0 [−0.06 – 0.05]	0 [−0.05 – 0.06]	0.01 [−0.02 – 0.04]	0.02 [−0.01 – 0.05]
Trend after PBF withdrawal										
Control group	1.03 [0.98 – 1.09]	0.98 [0.96 – 1]	0.97 [0.91 – 1.02]	1 [0.96 – 1.05]	1.03 [0.97 – 1.09]	1.06 [1.02 – 1.09]	0.94 [0.89 – 1]	0.95 [0.92 – 0.98]	0.99 [0.97 – 1.01]	1.01 [0.99 – 1.03]
Intervention group	0.98 [0.96 – 1]	1.04 [0.99 – 1.09]	0.99 [0.95 – 1.04]	0.96 [0.9 – 1.02]	1.05 [1.02 – 1.09]	1.04 [0.97 – 1.11]	0.95 [0.92 – 0.98]	0.94 [0.89 – 1]	1 [0.98 – 1.02]	0.98 [0.96 – 1.01]
Difference	0.05 [−0.01 – 0.11]	0.06 [0 – 0.11]	−0.03 [−0.1 – 0.04]	−0.04 [−0.11 – 0.03]	−0.02 [−0.09 – 0.05]	−0.02 [−0.09 – 0.06]	−0.01 [−0.07 – 0.06]	−0.01 [−0.07 – 0.05]	−0.01 [−0.04 – 0.02]	−0.02 [−0.05 – 0.01]

Simple multilevel represents the multilevel model adjusted for the time-invariant covariates. Multilevel with PSW represents the multilevel model adjusted for the time-invariant covariates and propensity score weights.

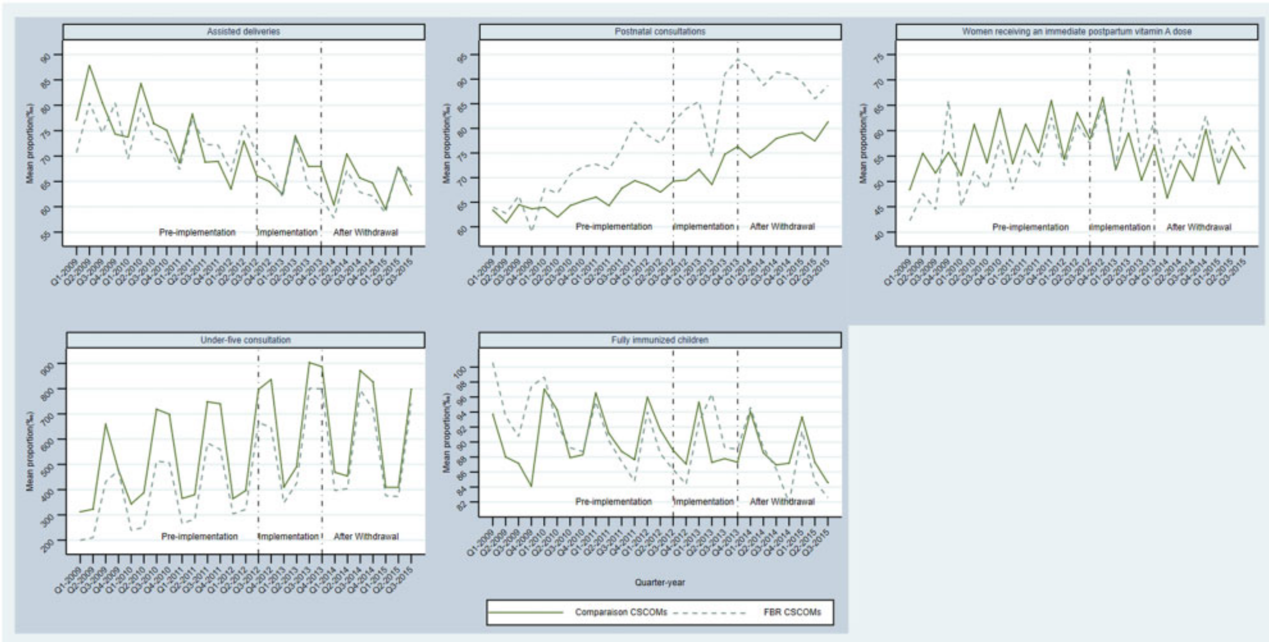

Figure 1 Evolution of maternal and child health in the pilot PBF CSCOMs ($n = 26$) and control CSCOMs ($n = 95$).

programme theory of change is essential for PBF schemes to ensure a well-thought-through programme design and evaluations that explain how and why changes occur (Ida and Bastøe, 2015).

Our finding that the introduction of the pilot PBF programme did not globally bear a significant impact on maternal and child health services utilization in the Koulikoro region of Mali is consistent with the findings from a recent study using a quasi-experimental design with an independent control group that did not clearly demonstrate the ineffectiveness of PBF in improving MCH indicators in Burkina Faso (Zizien *et al.*, 2018). A similar analysis of routine data was carried out in Benin for the 2010–15 period in 400 health facilities and showed the same lack of effect of the PBF programme on the use of maternal and child health services (Johnson *et al.*, 2016). That programme was organized over a longer period (4–6 years), but it showed notable improvements in the quality of care (Johnson *et al.*, 2016), which we have not been able to study in our research. The lack of effect on immunization coverage among children in the context of Benin was attributed to the already high level of this indicator, which left no room for significant improvement (Johnson *et al.*, 2016). Certainly, as the use of maternal and child health services was globally low in Mali (Ponsar *et al.*, 2011) and Benin (Johnson *et al.*, 2016), we expected a significant improvement even for a short duration of intervention, and, surprisingly, that was not the case. It may, therefore, be necessary to question the theory of the programme (Ida and Bastøe, 2015), the implementation factors and the causal hypothesis promoted by the PBF (Paul and Renmans, 2018).

In contrast, using data from the national health information system, another study concluded that the pilot PBF programme implemented in three districts in Burkina Faso has led, in a relatively short period, to increases in the numbers of consultations (27.7%), deliveries (9.2%) and advanced postnatal consultations (119%) (Steenland *et al.*, 2017). The very short-term effectiveness of the PBF intervention in Burkina Faso could be explained by the fact that the PBF programme operated for 18 months and that it was closely monitored by the World Bank and the Technical Secretariat for the programme implementation. This was not the case in Mali, where the programme implementation suffered from weaknesses that reduced the chance of the programme having a substantial effect on maternal and child healthcare indicators (Seppey *et al.*, 2017).

Methodological considerations

Our study builds on previous research and rigorously contributes to the ongoing debate on the effectiveness of PBF interventions using a quasi-experimental design and a mixed-effects negative binomial combined with linear spline regression model. These analytical strategies allowed us to choose the best-fitting model between a covariates adjustment and a propensity weighting approach, and have certainly contributed to reduce the confounding when using routine data to evaluate public health interventions (Lagarde, 2012). In addition, they have allowed us to better appreciate the trends before PBF, during its implementation and after its withdrawal.

However, our study has several limitations that must be considered when interpreting the results. In spite of our efforts to limit biases by combined control interrupted time series with covariates adjustment, the non-random selection of intervention (and hence also control) facilities represents an important limitation, as covariates adjustment can only account for confounding on observable variables (Austin, 2011). We cannot exclude that intervention and control facilities differed also on unobservable characteristics and hence that our effect estimation remains biased by potential confounding (Shadish *et al.*, 2002; Brookhart *et al.*, 2010). Moreover, relying on secondary HMIS data meant that we had no information on the socioeconomic and clinical characteristics of individual patients; hence, we could not exploit variation at the individual level to maximize the precision of our estimates by controlling for additional confounders. Last, we cannot exclude the presence of some measurement errors in the HMIS data, although the completeness rates ranged from 95% to 100%. Since we assume HMIS measurement errors to be constant across intervention and control facilities, however, we postulate that this type of error should not have any effect on nor invalidate our findings.

Moreover, as noted earlier, we could not include random slope effects in our model to examine the heterogeneity of the intervention effects beyond what we could do through a visual exploration of the data. Since field experience across African settings increasingly reports on differential effects of PBF across facilities, we identify this as a potential future area of research, possibly in settings with a larger number of intervention clusters and hence better statistical power to explore also effect heterogeneity. In our specific case, since the intervention was deployed sequentially and probably with varying intensity, it is plausible to assume that some CSCOMs might have responded more favourably than others to the intervention.

Conclusion

Implementation of the pilot PBF programme in the Koulikoro region of Mali has provided exceptional conditions for a timely assessment of its effects. Overall, the introduction and withdrawal of the pilot PBF programme did not have significant effects on the coverage of maternal and child health services in the Koulikoro region of Mali. Although several gains were generated by the implementation of the PBF pilot project in Mali according to its promoters (Toonen *et al.*, 2014), notably in terms of investments in human and material resources, the absence of significant effects in the whole CSCOMs could be explained by the context, by the weaknesses in the intervention design and by the causal hypothesis and implementation. Further qualitative inquiry is required in order to provide policy-makers and practitioners with vital information about the lack of effect detected by our quantitative analysis. Similarly, it would be of value also to study the impact of PBF on additional indicators, such as the quality of service delivery and the equality of access to maternal and child healthcare (Ridde *et al.*, 2018).

Supplementary data

Supplementary data are available at *Health Policy and Planning* online.

Acknowledgements

The authors thank the health authorities in Koulikoro region of Mali and the regional HMIS manager for facilitating data collection. A special thanks to Association Miseli members for logistic support for data collection. The authors would like to thank Salvador Shabbir for editing support.

This article is drawn from a research programme funded by the International Development Research Centre (IDRC) of Canada.

Conflict of interest statement. The authors do not have any conflict of interest.

Ethical approval. The research was accepted by the Ministry of Health and ethics committees in Mali and Canada (CRCHUM).

References

Abadie A, Diamond A, Hainmueller J. 2010. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *Journal of the American Statistical Association* **105**: 493–505.

Abdourahmane C, Tony Z, Christian D, Ridde V. 2018. La mise en œuvre du FBR dans les CSCom au Mali: quelles leçons retenir? *MISELI, Note de Politique*. Retrieved from http://www.documentation.ird.fr/hor/fdi:010076728, accessed 12 July 2018.

Akaike H. 1998. Information theory and an extension of the maximum likelihood principle. In: Parzen E, Tanabe K, Kitagawa G (eds). *Selected Papers of Hirotugu Akaike*. New York, NY: Springer New York, 199–213.

Austin PC. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research* **46**: 399–424.

Barnett AG, Dobson AJ. 2010. Controlling for season. In: Barnett AG, Dobson AJ (eds). *Analysing Seasonal Health Data*. Berlin, Heidelberg: Springer Berlin Heidelberg, 129–150.

Basinga P, Gertler P, Binagwaho A *et al*. 2011. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *The Lancet* **377**: 1421–8.

Binyaruka P, Patouillard E, Powell-Jackson T *et al*. 2015. Effect of paying for performance on utilisation, quality, and user costs of health services in Tanzania: a controlled before and after study. *PLoS One* **10**: e0135013.

Bonfrer I, Van de Poel E, Van Doorslaer E. 2014. The effects of performance incentives on the utilization and quality of maternal and child care in Burundi. *Social Science & Medicine* **123**: 96–104.

Brookhart MA, Sturmer T, Glynn RJ, Rassen J, Schneeweiss S. 2010. Confounding control in healthcare database research: challenges and potential approaches. *Medical Care* **48**: S114–120.

Coalition des Alternatives Africaines Dettes et Développement (CAD-Mali). 2011. Etude sur « l'état des lieux des sages-femmes au mali »: cas de la région de Koulikoro. Retrieved from https://www.cadmali.org/IMG/pdf/Rapport_de_l_Etude_sur_l_etat_des_lieux_des_Sages-Femmes_au_Mali.pdf, accessed 12 July 2018.

Das A, Gopalan SS, Chandramohan D. 2016. Effect of pay for performance to improve quality of maternal and child care in low- and middle-income countries: a systematic review. *BMC Public Health* **16**: 321.

Dimick JB, Ryan AM. 2014. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA* **312**: 2401–2.

D'Ostie-Racine L, Dagenais C, Ridde V. 2013. An evaluability assessment of a West Africa based non-governmental organization's (NGO) progressive evaluation strategy. *Evaluation and Program Planning* **36**: 71–9.

Eichler R. 2006. *Can "Pay for Performance" Increase Utilization by the Poor and Improve Quality of Health Services*. Washington, DC: Center for Global Development.

Feiveson AH. 1999. *FAQ: What Is the Delta Method and How Is It Used to Estimate the Standard Error of a Transformed Parameter*. Retrieved from: http://www.stata.com/support/faqs/statistics/delta-method/, accessed 25 May 2018.

Gautier L. 2016. Le financement basé sur les résultats au Mali. *Équité Santé*. https://www.researchgate.net/profile/Lara_Gautier/publication/297760179_Le_financement_base_sur_les_resultats_au_Mali/links/56e316f008ae65dd4cbac10f.pdf, accessed 9 October 2019.

Handley MA, Lyles CR, McCulloch C, Cattamanchi A. 2018. Selecting and improving quasi-experimental designs in effectiveness and implementation research. *Annual Review of Public Health* **39**: 5–25.

Heinmüller R, Dembélé YA, Jouquet G, Haddad S, Ridde V. 2012. Free healthcare provision with an NGO or by the Malian governmentimpact on health center attendance by children under five. *Field Actions Sci Rep*. Retrieved from: http://factsreports.revues.org/1731, accessed 9 December 2018.

Huillery E, Seban J. 2014. *Pay-for-Performance, Motivation and Final Output in the Health Sector: Experimental Evidence from the Democratic Republic of Congo*. Retrieved from: https://www.journals.uchicago.edu/doi/pdfplus/10.1086/703235, accessed 18 July 2018.

Ida L, Bastøe PØ. 2015. Results-based financing has potential but is not a silver bullet—theory-based evaluations and research can improve the evidence base for decision making. *The Evaluation Department, Norad*.

Johnson P, Bello K, Meessen B, Korachais C. 2016. Effets des expériences du financement basé sur les résultats (FBR) dans le secteur de la santé au Bénin. Une analyse à partir des données de routine collectées de 2010 à 2015. Antwerp, Belgium: IMT Anvers, 86.

Jung RC, Kukuk M, Liesenfeld R. 2006. Time series of count data: modeling, estimation and diagnostics. *Computational Statistics & Data Analysis* **51**: 2350–64.

Lagarde M. 2012. How to do (or not to do) . . . assessing the impact of a policy change with routine longitudinal data. *Health Policy and Planning* **27**: 76–83.

Lassi ZS, Kumar R, Bhutta ZA. 2016. Community-based care to improve maternal, newborn, and child health. In: Black RE, Laxminarayan R, Temmerman M, Walker N (eds). *Reproductive, Maternal, Newborn, and Child Health: Disease Control Priorities, Third Edition (Volume 2)*. Washington, DC: The International Bank for Reconstruction and Development/The World Bank.

Lawrence C. 2002. *Quantitative Applications in the Social Sciences: Spline Regression Models*. Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781412985901.

Lee J, Little TD. 2017. A practical guide to propensity score analysis for applied clinical research. *Behaviour Research and Therapy* **98**: 76–90.

Linden A, Adams JL. 2010. Evaluating health management programmes over time: application of propensity score-based weighting to longitudinal data. **16**: 180–5.

Lopez Bernal J, Cummins S, Gasparrini A. 2018. The use of controls in interrupted time series studies of public health interventions. *International Journal of Epidemiology* **47**: 2082–93.

McCaffrey DF, Ridgeway G, Morral AR. 2004. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods* **9**: 403–25.

Paul E, Albert L, Bisala BNS *et al*. 2018. Performance-based financing in low-income and middle-income countries: isn't it time for a rethink? *BMJ Global Health* **3**: e000664.

Paul E, Renmans D. 2018. Performance-based financing in the heath sector in low- and middle-income countries: is there anything whereof it may be said, see, this is new? *The International Journal of Health Planning and Management* **33**: 51–66.

Ponsar F, Van Herp M, Zachariah R *et al*. 2011. Abolishing user fees for children and pregnant women trebled uptake of malaria-related interventions in Kangaba, Mali. Health Policy. *Health Policy and Planning* **26**: ii72–83.

Renmans D, Holvoet N, Criel B, Meessen B. 2017. Performance-based financing: the same is different. *Health Policy and Planning* **32**: 860–8.

Renmans D, Holvoet N, Orach CG, Criel B. 2016. Opening the 'black box'of performance-based financing in low-and lower middle-income countries: a review of the literature. *Health Policy and Planning* **31**: 1297–309.

Ridde V, Gautier L, Turcotte-Tremblay AM, Sieleunou I, Paul E. 2018. Performance-based financing in Africa: time to test measures for equity. *International Journal of Health Services* **48**: 549–61.

Rosenbaum PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**: 41–55.

Seppey M, Ridde V, Touré L, Coulibaly A. 2017. Donor-funded project's sustainability assessment: a qualitative case study of a results-based financing pilot in Koulikoro region, Mali. *Globalization and Health* **13**: 86.

Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi?Experimental Designs: For Generalized Casual Inference*. Boston, MA: Houghton, Mifflin and Company.

Ssengooba F, McPake B, Palmer N. 2012. Why performance-based contracting failed in Uganda–an "open-box" evaluation of a complex health system intervention. *Social Science & Medicine* **75**: 377–83.

Steenland M, Robyn PJ, Compaore P *et al*. 2017. Performance-based financing to increase utilization of maternal health services: evidence from Burkina Faso. *SSM—Population Health* **3**: 179–84.

Toonen J, Dao D, Matthijssen J, Koné B. 2014. Evaluation finale: Accélérer l'atteinte de l'OMD 5 dans la région de Koulikoro—Projet pilote financement basé sur les résultats dans les cercles de Dioïla et Banamba. *Royal Tropical Institute KIT Development Policy & Practice*. Retrieved from: https://www.kit.nl/publications/page/2/, accessed 12 July 2018.

Wagenaar BH, Sherr K, Fernandes Q, Wagenaar AC. 2015. Using routine health information systems for well-designed health evaluations in low- and middle-income countries. *Health Policy and Planning*

Wholey JS. 2004. Evaluability assessment. *Handbook of Practical Program Evaluation* **2**: 33–62.

Witter S, Fretheim A, Kessy FL, Lindahl AK. 2012. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database of Systematic Reviews*. doi: 10.1002/14651858.CD007899.pub2.

Wiysonge CS, Paulsen E, Lewin S *et al*. 2017. Financial arrangements for health systems in low-income countries: an overview of systematic reviews. *Cochrane Database of Systematic Reviews*. doi: 10.1002/14651858.CD011084.pub2.

World Health Organization. 2014. *WHO Global Health Expenditure Atlas*. https://www.who.int/health-accounts/documentation/atlas.pdf, accessed 22 September 2018.

Zeng W, Shepard DS, Rusatira JdD Blaakman AP, Nsitou BM. 2018. Evaluation of results-based financing in the Republic of the Congo: a comparison group pre–post study. *Health Policy and Planning* **33**: 392–400.

Zizien ZR, Korachais C, Compaoré P, Ridde V, De Brouwere V. 2018. Contribution of the results-based financing strategy to improving maternal and child health indicators in Burkina Faso. *The International Journal of Health Planning and Management* **34**: 111–29.

Zombré D, De Allegri M, Ridde V. 2017. Medicine: immediate and sustained effects of user fee exemption on healthcare utilization among children under five in Burkina Faso: a controlled interrupted time-series analysis. **179**: 27–35.